

On Entropies in the Classification of Commodities – a Challenge for E-Commerce

Peter J. A. Reusch, L. Esmeralda Garcia Moreno

Summary

In classification systems like eCI@ss the number of classes grows and “everywhere” products like f.ex. screws are spread ambiguously across several segments and many classes. Entropy functions on test sets are used to calculate the dispersion of “everywhere” products into many classes across the classification system. For such dispersed classes the corresponding property sets are compared and it is shown that there is a real need for the harmonisation of property sets i.e. to make classification structures unambiguous.

1. Introduction

Manufacturers and distributors classify their products for e-commerce purposes so that potential customers can find these products in electronic markets. In B2B applications the customer is also interested in the classification of the commodities, because here the business customer wants to inform many sellers about his demands. The customer wants to get quotations and he wants to compare offers. For many products the comparison and evaluation of offers should mainly be done by computers. Therefore we need classified products with unambiguous unique and completely determined properties. If so the computer can help to compare and evaluate offers.

In the classification system eCI@ss for example today (version 6.0) there are 32590 classes in a 4-level hierarchy identified by a numeric code and 10930 attributes for the description of the commodities in these classes. Like the classes the attributes are also identified by a numeric code. The number of classes and attributes have grown from release to release. And the more classes there are the harder it is to detect the right class of a product – since from many points of view classes seem to overlap and their boundaries become fuzzy.

When systems for the classification of commodities were initiated - years ago - at the very beginning it was easy to establish commodity classes for example for a screw in a higher level class of machine elements for fixing and mounting (in eCI@ss: 23-11 in top segment 23).

In the last years in classification systems like eCI@ss the commodity classes were more and more differentiated and subdivided in order to specify the usage of commodities. A screw to fix a painting at the wall is different from the screw the surgeon needs to fix broken bones in a human body. The surgeon needs a screw made of materials that do not endanger human health and safety.

The table below shows that “everywhere” materials like pipes and pumps are described in 11 out of 26 main segments of eCI@ss. And there are many more common products which are classified in several segments across the eCI@ss structure like f.ex. screws, fittings, wheels, shafts, etc.. .

Examples of everywhere products classified ambiguously into several segments of eCI@ss

		eCI@ss top segment	Pipe	Pump	Fitting	Spring	Wheel	Shaft	Spindle
1	16	Food							
2	17	Machine device (spec. Appl.)		x					
3	18	Equip. mining, ...				x		x	
4	19	Information, comm.	x				x		
5	20	Packing	x	x	x				
6	21	Manufacturing	x	x	x	x	x	x	
7	22	Construction	x	x	x			x	
8	23	Machine element	x		x	x	x	x	x
9	24	Office product		x		x	x		x
10	25	Service							
11	26	Energy							
12	27	Electric engin.	x	x	x	(x)	x	x	x
13	28	Automotive techn.			x		x		
14	29	Home economics, techn.							
15	30	Aux. Supply		x					
16	31	Polymers							
17	32	Laboratory material, techn.		x	x			x	
18	33	Installation	x						
19	34	Medicine		x				x	
20	35	Semifinished products	x						
21	36	Machine apparatus	x	x		x	x	(x)	x
22	37	Industrial piping	x		x				
23	38	Inorganic chem.							
24	39	Organic chem.							
25	40	Occupational safety	x	x	x		x		
26	41	Marketing			x				
			11/26	11/26	10/26	5/26	8/26	7/26	4/26

Table 1: Examples of everywhere products classified ambiguously into several segments of eCI@ss

2. Entropies

A pipe is a “everywhere” product type – and most people will have a core concept of a pipe in their mind. But “pipes” are no longer referenced to a single product class in the classification system like eCI@ss. The number of top segments and the number of classes that share such an “everywhere” product type is an indication of entropy of a classification system.

In information theory¹ the entropy of a discrete random variable X with possible values $\{x_1, \dots, x_n\}$ is

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

Here p denotes the probability function of X .

¹ Claude E. Shannon, Warren Weaver: : The Mathematical Theory of Communication. Univ of Illinois Press, 1949.

An entropy function to characterise a classification system can be developed starting with a set of test objects (TO) – for example a set $TO = \{ \text{pipe, screw, ...} \}$ – and with the number of segments or classes of the classification system which hit these test objects divided by the total number of segments or classes of the classification system. In the table above we have

a given set of segments = $GS = \{ x_1=16, x_2=17, \dots, x_{26}=41 \}$

and a relative number of hits of the test objects in these segments

$$HO = \{ p(x_1)=11/26, \dots \}.$$

According to this approach the entropy of the pipe as a test object for the 26 main segments is

$$H(\text{pipe} \mid \text{main segments of eCl@ss}) = - (11/26) \log_2(11/26) = 0,5250419$$

and for the full test set we apply

$$H(X) = - \sum_{x_i \in TO} p(x_i) \log_2 p(x_i)$$

The following figure shows the entropy of “screw” in all segments and classes.

Entropy of the term "screw" _in segments and classes of the eCl@ss structure

			6.0
		eCl@ss top segment	English
			Screw
1	16	Food	
2	17	Machine device (spec. Appl.)	
3	18	Equip. mining, ...	*
4	19	Information, comm.	
5	20	Packing	*
6	21	Manufacturing	1
7	22	Construction	1
8	23	Machine element	1
9	24	Office product	
10	25	Service	
11	26	Energy	
12	27	Electric engin.	1
13	28	Automotive techn.	
14	29	Home economics, techn.	
15	30	Aux. Supply	
16	31	Polymers	
17	32	Laboratory material, techn.	*
18	33	Installation	
19	34	Medicine	1
20	35	Semifinished products	
21	36	Machine apparatus	*
22	37	Industrial piping	*
23	38	Inorganic chem.	
24	39	Organic chem.	
25	40	Occupational safety	
26	41	Marketing	
			5/26
		count/total	0,19
		entropy of segment	0,46

Entropy of classes

	6.0
	English
	Screw
Total	32593
Selected	51
count/total	0,00156475
Entropy of classes	0,01458326

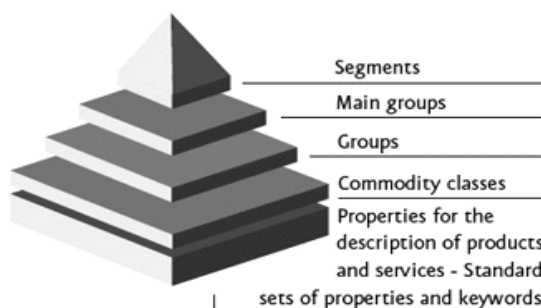


Table 2 : Entropy of the term "screw" _in segments and classes of the eCl@ss structure

The following figure shows the entropy for the term “wheel” and the values of the entropy function for eCI@ss in the English version 5.1.1 and in the corresponding Spanish version. In the Spanish version the term of “wheel” has to be translated either as to “volante” or to “rueda”. Consequently entropy may vary across the languages.

Entropies of classes

	5.1.1	5.1.1	5.1.1
	English	Spanish	Spanish
	Wheel	Volante	Rueda
Total	27216	27216	27216
Selected	17	1	19
count/total	0,00062	3,7E-05	0,0007
Entropies of classes	0,00665	0,00054	0,00732

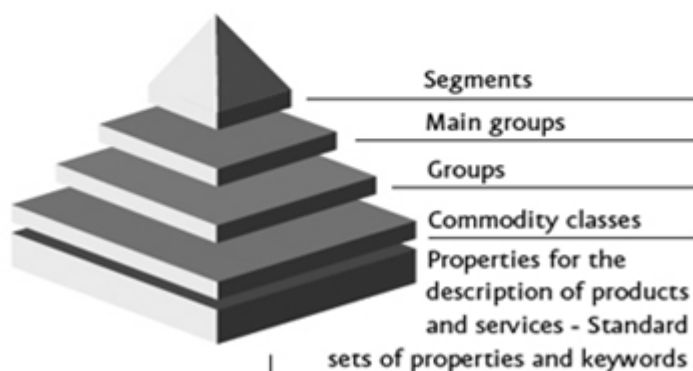


Table 3a : Entropy of the term “wheel” in eCI@ss segments and classes in English and Spanish language version

Entropies in the classification of commodities

			5.1.1	5.1.1	5.1.1
			English	Spanish	Spanish
			Wheel	Volante	Rueda
		eCl@ss top segment			
1	17	Machine device (spec. Appl.)			
2	18	Equip. mining, ...			
3	19	Information, comm.	*	*	
4	20	Packing			
5	21	Manufacturing	1		1
6	22	Construction			
7	23	Machine element	1		1
8	24	Office product	1	*	1
9	25	Service			
10	26	Energy			
11	27	Electric engin.			1
12	28	Automotive techn.	1		1
13	29	Home economics, techn.			
14	30	Aux. Supply			
15	31	Polymers			
16	32	Laboratory material, techn.			*
17	33	Installation			*
18	34	Medicine			
19	35	Semifinished products			
20	36	Machine apparatus	*	*	*
21	37	Industrial piping			
22	38	Inorganic chem.			
23	39	Organic chem.			
24	40	Occupational safety	1	1	
25	41	Marketing		*	
			5/25	1/25	5/25
		count/total	0,20	0,04	0,20
		entropies in segments	0,46	0,19	0,46

Table 3b : Entropy of the term “wheel” in eCl@ss segments and classes from language versions in English and Spanish

This approach can only be applied to get an indication for the growing “chaos” in a classification system – growing from level to level. This indication can be used to initiate special control procedures – for example a control of the attributes of corresponding classes in various segments of the classification system. The attribute list of corresponding classes can be compared in a computer assisted way.

The results of the entropy functions highly depend on preselected sets of test objects. Here expert knowledge is needed to build appropriate sets of objects with a focus on critical segments of a classification system. We do not need test objects for the chemicals in segment 38 or 39 of eCl@ss that much, since such chemicals will not appear in other segments of the classification system in the next years.

The authors tested about 100 test objects and a reasonable number of test sets – and in most cases the results were useful to focus on critical classes – especially to start the harmonisation of properties in critical classes.

3. Property Sets Today

The following table 4 shows a list of properties used for the description of screws. 51 classes in eCI@ss 6.0 are dealing about screws – 5 of them are shown in the columns of the table.

The properties assigned to these classes are different. That is not a main problem, since to identify a screw for the surgeon – shown in the last column – we need properties (like the “Pharmaceutical central number”) – which are not used for other screw descriptions.

For other medical products there are far more specific properties – like for the heart valve description– class 34-32-02-02 :

- AAA300001 - actuarial durability of heart valve prosthesis at 10 years
- AAA301001 - actuarial durability of heart valve prosthesis at 15 years
- AAA302001 - actuarial durability of heart valve prosthesis at 20 years
- AAA305001 - allergenic potential
- AAA315001 - Anticalcification of bioprosthesis
- AAB058001 - designated sterilization method
- AAA567001 - Developer or inventor of heart valve prosthesis
- AAA664001 - Diameter of tissue annular ring
- AAA554001 - effective orifice area of heart valve prosthesis in vitro
- AAA555001 - effective orifice area of heart valve prosthesis in vivo
- AAB056001 - estimated durability of heart valve prosthesis
- BAJ013001 - Health Industry Barcode
- BAA001002 - Manufacturer name
- BAD847002 - Manufacturer product number
- AAA863001 - Material of intracardiac or vascular implant
- AAA956001 - MRI-compatible
- BAJ012002 - Pharmaceutical central number
- AAA985001 - Position of heart valve prosthesis
- BAA316002 - Product name
- BAA002002 - Product type description
- AAA310001 - Rate of thrombo-embolic and haemorrhagic occurrences
- AAA487001 - Reference norm product
- BAB542001 - Supplier name
- BAA059002 - Supplier product number
- AAA378001 - Type of heart valve prosthesis

List 1: Properties of class 34-32-02-02 Heart valve in eCI@ss 6.0

Regarding the entirety of screws – it is surprising that for some screws – used by surgeons – fundamental attributes like “Screw length” are missing. Such fundamental should be assigned to all classes dealing with screws – and there are 51 of such classes.

When we comparing the property sets of classes dealing about screws we can easily capture from table 4 that there are no columns which do have an identical set of properties.

Property set	Hexagon head cap screw	Special screw	Half-counter sunk screw	Wood screw	Cortical screw (osteo-synthesis)
	23-11-01-01	23-11-01-10	23-11-01-21	23-11-01-11	34-32-17-01
BAA271002 - EAN code	x	x	x	x	
BAA898001 - Drive quantity	x			x	
BAJ013001 - Health Industry Barcode					x
BAA001002 - Manufacturer name	x	x	x	x	x
BAA002002 - Product type description	x	x	x	x	x
BAA059002 - Supplier product number	x	x	x	x	x
BAD847002 - Manufacturer product number	x	x	x	x	x
BAA316002 - Product name	x	x	x	x	x
BAA899001 - Order supplement code	x	x		x	
BAA900001 - Order supplement according to standard	x	x		x	
BAA907001 - Thread diameter	x	x		x	
BAA909003 - Thread pitch at nut end	x	x		x	
BAA916001 - Head diameter of screw	x	x		x	
BAA917001 - Head form	x	x		x	
BAA918001 - Position of thread coating	x	x		x	
BAA919001 - Screw length	x	x		x	
BAA922001 - Length of thread coating	x	x		x	
BAA929001 - Product class in accordance with norm	x	x		x	
BAA932001 - Key width	x			x	
BAA936001 - Tolerance information in accord. with norm	x	x		x	
BAA997001 - Thread length	x	x		x	
BAB010001 - Publication date (year-month)	x	x		x	
BAB072002 - Tolerance	x	x		x	
BAB090001 - Head form pursuant to standard	x	x		x	
BAB101001 - Surface protection	x	x		x	
BAB112001 - Material in accordance with norm	x	x		x	
BAB150001 - Surface protection in accord. with norm	x	x		x	
BAB162001 - Height of head	x	x		x	
BAB165002 - standard letter to the standard number	x	x		x	
BAB341001 - Thread size	x	x			
BAB618001 - Thread design according to standard letter	x	x		x	
BAB637001 - Product class	x	x		x	
BAB664003 - Material	x	x		x	
BAE162001 - Requirement in accordance with	x	x		x	
BAA914001 - Width of screw head		x		x	
BAA930001 - Shaft diameter of the screw		x			
BAB093001 - Head length of screw		x		x	
BAB351001 - Metric connecting thread size					
BAC680002 - Max. cross section					
BAC745002 - Min. cross section					
BAC870001 - Condition of surface					
BAB542001 - Supplier name					x
BAJ012002 - Pharmaceutical central number					x
TOTAL	33	34	6	34	8

Table 4 : Properties used in selected classes dealing about screws.

4. Proposal

As shown in this paper there is a real need to harmonise the property sets of classes especially for “anywhere” products - and to start with this as soon as possible.

We need hierarchical property sets – starting with the fundamental properties for all corresponding classes and going on with additional specific properties.

Property sets must be harmonised before new data models – ontologies – for eCI@ss can be introduced.

References:

Claude E. Shannon, Warren Weaver: The Mathematical Theory of Communication. Univ of Illinois Press, 1949. ISBN 0-252-72548-4

Authors:

Prof. Dr. Peter J. A. Reusch
L. Esmeralda Garcia Moreno

Fachhochschule Dortmund
Fachbereich Wirtschaft
Emil-Figge-Str. 44
44227 Dortmund
Germany

Peter.Reusch@fh-dortmund.de
esmeraldagar56@gmail.com

The authors thank Dr. Thomas Holzapfel for supporting this paper.

The project thesis and the master thesis of Laura Esmeralda Garcia Moreno in the European Master in Project Management – EuroMPM – at the University of Applied Sciences and Arts in Dortmund show details of this research.